



September 26, 2000

**DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM
SERIES B-16**

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Michael J. Batutis *MJB*
Chief, Decennial Full Count Review

Prepared by: Gretchen Stiers
Decennial Full Count Review

Subject: Demographic Full Count Review Report

The attached document is a prototype of the report that we will prepare, per your request, following completion of applicable Census operations. The completed report is intended to aid the Executive Steering Committee on A.C.E. Policy (ESCAP) in its recommendation regarding the release of the statistically corrected data or the data without statistical correction as the P.L. 94-171 data. This report, together with other reports, will assess the operations and results of both the initial Census and the A.C.E. Both sets of assessments will be available to the ESCAP to aid the Committee in reaching its recommendation regarding the use of the statistically corrected data.

The attached prototype contains both empty table shells and a description of textual analysis that will assess specific aspects of the applicable operations. This report focuses on the results of the Demographic Full Count Review undertaken by the Population Division and the Housing and Household Economic Statistics Division. It will discuss output measures related to the review of 100% census data items, and will identify issues and patterns related to data quality.

It is important to note that the conduct of the operations may lead us to modify the attached format by including additional information. It is also likely that descriptions and definitions will be enhanced or the data items could undergo revision. Conversely, we may conclude, for a variety of reasons, that some of the information set forth in the attached prototype may not be available. The attached document sets forth our conclusions prior to completion of the A.C.E. about what information would properly inform the ESCAP on this subject, but is subject to modification.

Census 2000 Demographic Full Count Review Report

Prepared by Gretchen Stiers

Introduction

Demographic Full Count Review draws on comparative demographic data and expertise to rapidly examine, rectify if possible, and clear Census 2000 files and products for subsequent processing or for release to the public. The review checks for data reasonableness, internal and inter-product consistency, and consistency with historical and external data sources. The objective of the review is to identify, investigate, and document issues with the data. The review strategy calls for capitalizing on every opportunity to improve the data. It is an iterative process, where insight into data quality builds through reviewing files generated in each processing cycle and culminates in clearance of redistricting data as required under PL 94-171.

Demographic Full Count Review encompasses all Census 2000 data files and products. For the purposes of this report, however, the discussion is limited to the review of 100% data items. That is, the responses to the questions that appear on both the short form and long (sample) form. Subsequent analysis will address the data that is collected from long form submissions. The 100% data include:

- Total population
- Group quarters population
- Group quarters units
- Housing units
- Household population
- Age
- Race
- Sex
- Ethnic and Hispanic origin
- Relationship
- Tenure

The Count Review analysis serves two principal purposes in census evaluation:

1. The review process supports Census management in deciding whether to clear files for subsequent stages of processing. Census files are created state-by-state and each state's data passes through several processing stages:
 - a. HCUF – Hundred percent Census Unedited File
 - b. HCEF – Hundred percent Census Edited File
 - c. HDF – Hundred percent Detail File
 - d. HEDF – Hundred percent Estimated Detail File
 - e. Redistricting Products

Analysts will review the data file for each state at each stage, building an increasingly complete picture of data quality. Starting with the HCEF, each file is cleared by the Population Division prior to the file being released for the next stage (or release to the public in the case of the redistricting products). The Count Review process is designed to support the clearance decision.

2. At the completion of the 100% review for all states, the count review process results will provide a comprehensive view of data quality for the 100% data. The data will have been reviewed to low (tract and place) levels of geography to identify, investigate and document any issues that exist in the data. This insight will support subsequent consideration of count questions raised by localities and will serve as a “lessons learned” resource for future censuses.

Methodology

The demographic full count review checks for consistency of Census 2000 data with historical and external data sources. This is accomplished by applying a team of analysts with demographic expertise spanning the 100% data variables. In addition, Federal State Cooperative Program for Population Estimates (FSCPE) representatives will assist the Census Bureau in conducting the review. The analysts are supported with effective automated tools so they can rapidly identify issues, efficiently investigate causes, and quickly document their findings.

Comparative Data

The comparative data used in the review allow the analyst to rapidly identify inconsistencies, outliers, or other anomalies in the Census data. The analysts have developed specifications for analytic applications that will facilitate the comparisons against benchmark data. These will be developed using a variety of tools ranging from creation of simple difference tables to complex presentations using a geographic information system.

Comparative data that comprise the core benchmark database include:

- **1990 Census data.** This will include 1990 data retabulated using the latest available geography.
- **Population Division Estimates data.** The 1999 estimates data extrapolated to April 1, 2000. For comparisons at the place level, the data will be tabulated using the latest available geography.
- **Claritas data.** Independent, commercially available population data.

Additional comparative data may be used on a case-by-case basis to meet specific analytic objectives. These data would not be included in the core benchmark database and would not be used in programmed applications. They may, however, be a supplemental resource to be drawn

on during the review. The primary source of additional data will be comparative data drawn from local sources provided by the state representatives.

Analytic Perspectives

Subject Matter Experts. Analysts from the Population Division and the Housing and Household Economic Statistics Division will review the files. The various branches within these divisions have resident demographic expertise that spans the 100% data items (age, race, sex, etc.).

State Representatives. The Federal State Cooperative Program for Population Estimates (FSCPE) will support the Census Bureau in conducting the review. The FSCPE has a history of working with the Population Division on estimates and comparative data and have directly relevant technical expertise. The FSCPE participants will work as on-site analytic partners for technical reviews. They will use the same methods and tools as Census in-house analysts for those reviews, supplemented in some cases by specific analyses using locally developed data. In addition to the on-site work, the FSCPE participants will provide consultative support as “on call” local experts to apply expertise to specific questions referred to them.

FSCPE representatives will review Census 2000 data for their own state as a contractor with special sworn status and will work at the Census Bureau headquarters for all stages of the review. The person is an employee of an agency designated as the FSCPE organization for their state. The FSCPE representatives have taken an oath under Title 13, U.S. code Section 9, not to disclose any Title 13 confidential information, and not to disclose any other information about Census 2000 data or products prior to the official release of the redistricting data for their state. These requirements are based upon the need to assure the confidentiality of individual household data and to preserve the integrity of the data used for redistricting.

Analytic Tools

Given the pace and volume of data processing, the review will often require clearing files for two or more states per day. To manage this pace of analysis, the analysts will be supported with a suite of automated tools including SAS multi-dimensional database tools and a geographic information system. The analytic tools will facilitate:

- Comparison of Census 2000 data with benchmark data
- Analysis by characteristics – age, race, etc, including comparing Census 2000 data with benchmark data and displaying percent difference by variable (age, race, etc.).
- Drilling-down to the microdata in support of analytic investigation of problems

Analytic Strategy

The general approach to the review of census files is as follows:

1. Using the GIS, look for outlier counties based on comparing total population from 2000 data with 1990 and Estimates data.
2. Select an outlier county. Display tract data and look for the outlier tracts. Overlay place boundaries to get a sense for the population centers. Attempt to localize the problem.

3. Examine other variables (group quarters population, housing units). Look for underlying clues as to the population variation.
4. Shift to the SAS MDDB comparative tables. Find the county/tract and examine the tabular data for clues on the population disparity. Use the drilldown capability to check detailed data.
5. Try to determine the nature of the problem. Does it seem like a processing problem? An under/overcount of a specific locality? A geographic misallocation? What is the analyst's judgment of the root cause?
6. Assess the implications of the problem. Is a large population affected? Is the problem likely to crop up elsewhere (e.g., a processing problem)?
7. Document the findings and assessments.
8. Move on to the next outlier.

Subject matter analysts investigating characteristic data will apply a similar approach but will be examining data at the state and county level. They will be looking for broader demographic trends and reasonableness. Subject-specific tables will be generated to assist these analysts in their review. In addition, many analysts will focus on the impact of edit and imputation. Specialized tools have been developed to support this review as well.

Complementary Review – Demographic Analysis

The Demographic Analysis (DA) will be conducted in parallel with count review. DA is an analytic approach which has been extensively used at the Census Bureau to measure coverage of the national population in every census since 1960 (see Siegel and Zelnik, 1966; U.S. Bureau of the Census, 1974, 1988; and Robinson et al, 1993 for the basic demographic evaluations of the 1960 - 1990 censuses).

Demographic Analysis represents a macro-level approach, where analytic estimates of net undercount are derived by comparing aggregate sets of data or counts. The traditional DA population benchmarks are developed for the census date by analysis of various types of demographic data essentially independent of the census, such as administrative statistics on births, deaths, immigration, and Medicare enrollments, as well as estimates of emigration and undocumented immigration.

The Demographic Analysis complements the Demographic Full Count Review in that it is based on an independent set of benchmark data and is a powerful technique for detecting broad changes or patterns. By comparison, the count review analysis is carried to a finer degree of geographic granularity and facilitates more precise examination of individual characteristic variables (age, race, sex, etc.). The Demographic Analysis may well provide broad direction to

the count review analysts; the count review results may shed light on the detailed anomalies and root causes underlying broad patterns.

Review Limitations

The principle limitations of the count review approach are:

Time constraints. The pace of the review is dictated by the file-processing schedule. Historically, analysts have only been afforded a day to review a state file. Even with the power of the automated analytic tools it will be difficult to accomplish a comprehensive review in the time available.

Comparative data. The review uses historical census data and census-derived estimates for comparisons. The data are often not in exact one-to-one correspondence with the Census 2000 data at low levels of geography. Also, the analyst must separate significant differences from those which would be expected. While there are mitigating strategies for both of these difficulties, they nevertheless will be factors in conducting the review.

Results

The results from count review will serve both purposes described above. Analysts will identify and document issues. Management will use this documentation to support the clearance decision-making process. The issues will be captured in a database, creating a resource of insights that will grow in depth and breadth as the review progresses. Supplementing the state-by-state review, a review team will periodically review the data for emerging patterns or issues of broad implications. The results of this review will be documented in formal Count Review Reports as described below.

Analytic Output

Tract Results for Population Measures.

The analysis will support summarizing difference data for principle population measures by tract:

- Total population
- Group quarters population
- Group quarters units
- Housing units
- Household population

Table 1 shows the layout for a tabulation of total population differences by tract showing differences of Census 2000 data as compared to benchmark data. Similar tabular displays will be constructed for each of the major population variables.

Table 1. Total Population: Distribution Of Tracts By Percentage Difference

Percent Difference	Census 2000 vs. 1990 Census	Census 2000 vs. Census Estimates	Census 2000 vs. Claritas
<-25			
-25 - -20			
-20 - -15			
-15 - -10			
-10 - -5			
-5 - 0			
0 - 5			
5 - 10			
10 - 15			
15 - 20			
20 - 25			
>25			
Total Tracts			

Patterns that emerge from examining this data will be discussed and insights gained in the review as to possible root causes will be documented.

Results by Characteristic.

In general, the analysis of characteristic data will be carried to the county level. This includes:

- Age
- Race
- Sex
- Ethnic and Hispanic origin
- Tenure
- Relationship

Table 2 illustrates the presentation of findings from reviewing these characteristics. In this example, age range differences will be displayed for each state.

Table 2. Age Distribution: Percent Differences

Age Range	% Change from 1990 Census	% Change from Census Estimates

Similar tables will summarize findings for each of the characteristic variables. Patterns that emerge from examining this data will be discussed and insights gained in the review as to possible root causes will be documented.

Planned Review Phases

The review will commence with the release of the initial state HCUF file and will continue through the release of the redistricting data. To provide insight into the results of the review, two reports are planned. The Phase 1 report will cover the findings through the HDF review and apportionment release. The Phase 2 report will address the review results through the release of redistricting products, including the effect of the A.C.E. results.